

# 网格技术的发展及其应用研究<sup>y</sup>

史文<sup>明</sup> 曾文华

(厦门大学软件学院 厦门 361005)

**摘 要** 网格计算是一个新出现的名词,也是近年来兴起的一个研究热点,它是以大规模的资源协作共享、创新的应用以及高性能计算为特点,诞生的一个全新领域。本文在阅读大量最新的国内外相关文献后,对网格计算的概念、特点、基本体系结构进行简要介绍,并详细分析了网格的研究现状、应用领域以及发展趋势。

**关键词** 网格 网格计算 网格系统 OGSA 网格应用

**中图分类号** TP393

## Research of the Development of Grid Technology and Its Application

Shi Wenchong Zeng Wenhua

(Software School, Xiamen University, Xiamen 361005)

**Abstract** Grid computing has emerged as a new hot name and research field. It focuses on large-scale resource sharing, innovative applications, and in some cases, high-performance orientation. This article defines the grid computing, grid performance and architecture, and describes the current research on the grid, the applications of the grid, and the development trends of the grid.

**Key words** Grid, Grid computing, Grid system, OGSA, Grid application

**Class number** TP393

## 1 引言

在 1969 年,美国国防部利用数据包交换技术和开放式体系结构,组建了 ARPANET,这是因特网的前驱;1990 年,欧洲原子核研究中心(CERN),为了高能物理研究的需要,发明了万维网(WWW);而从 1993 年开始,高性能计算技术和互联网技术进一步融合,产生了网格(Grid)。网格技术旨在使互连网上所有资源,包括计算资源、存储资源、通信资源、软件资源、信息资源、知识资源等,实现全面共享与协同工作,使整个 Internet 整合成一台巨大的超级计算机,为用户提供“即连即用”式的服务。

### 1.1 网格的定义

对一种新兴事物的命名或定义经常颇费周折,网格也不例外。关于什么是网格和什么是网格计算,现在还没有一个统一的定义。到目前为止,有许多研究网格的学者对网格下过各种各样的定义。Ian Foster 在 2002 年 7 月从三方面对网格进行了清晰的描述<sup>[1,3,6]</sup>:

(1) 协调不服从集中式控制的资源。即网格能集成和协调资源与用户在不同控制域内的活动。

(2) 使用标准的、开放的、通用的协议和接口。即一个网格是由多用途协议和接口来构建的,该协议将能解决诸如鉴别、授权、资源发现和资源访问等一些基本问题。

(3) 提供非常的服务质量(QoS)。即允许按协作方式来使用其成分资源以提供各式服务质量。

### 1.2 网格的特点<sup>[9]</sup>

(1) 异构性(Heterogeneous): 网格系统由分布在 Internet 上的各类资源组成,这些资源是异构的,即它们的计算机体系结构、操作系统、网络布局等可能存在着差异,如何实现异构机器或机群之间的协作和转换是网格计算的首要问题。

(2) 可扩展性(Scalability): 即在计算规模不断扩大的情况下不会导致网格计算性能的降低。

(3) 动态自适应性(Adaptability): 即在网格计算中,某个或某些资源出现故障或问题,不应该影响整个服务的质量,因此需要动态监视和管理网格资

<sup>y</sup> 收到本文时间:2005 年 10 月 24 日

源。

2 网格系统的体系结构

2.1 网格体系结构的定义

文献<sup>[3]</sup>中定义网格体系结构就是关于如何建造网格的技术。它给出了网格的基本组成与功能,描述了网格各组成部分的关系以及它们集成的方式或方法,刻画了支持网格有效运转的机制。而在文献<sup>[1]</sup>中,网格体系结构是划分系统基本组件,指定系统组件的目的与功能,说明组件之间如何相互作用的技术。这两个对网格体系结构的基本定义略有不同,第一个定义强调的是组件与组件之间的相互作用;而第二个定义还要求说明各个部分如何集成为一个整体,以及通过何种机制实现整体的功能<sup>[3]</sup>。

准确定义网格的体系结构是构造网格的一个基础工作。到目前为止,比较重要的网格体系结构主要有两个,五层沙漏模型和开放网格服务结构(OGSA)。

2.2 五层沙漏模型

五层沙漏模型是 Ian Foster 等在早些时候提出的体系结构。它的基本思想和构架类似于计算机网络中的 TCP/IP 协议的构架。图 1 是五层沙漏结构与 TCP/IP 网络协议的一个对比图。

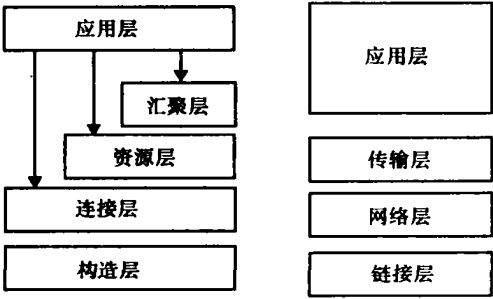


图 1 五层沙漏结构与 TCP/IP 网络协议的对比

- (1) 构造层 (Fabric Layer): 向上提供网格中可供共享的资源,它们是物理或逻辑的实体。
- (2) 连接层 (Connectivity Layer): 它是网格中网络事务处理通信与授权控制的核心协议。
- (3) 资源层 (Resource Layer): 对单个资源实施控制,与可用资源进行安全握手、对资源做初始化、监测资源运行状况、统计与付费相关的资源使用数据。
- (4) 汇聚层 (Collective Layer): 汇集资源层提交的受控资源,供虚拟组织的应用程序共享。
- (5) 应用层 (Application Layer): 这层是网格上

用户的应用程序。

五层模型中,每层中协议数量都不同。对于最核心部分,既要实现上层协议向核心协议映射,又实现核心协议向下层协议映射;核心协议应在所有支持网格计算的地点都得到支持,因此数量不能太多,这样核心协议就形成了协议层次结构中的一个瓶颈。在五层结构中,资源层和连接层共同组成这一核心的瓶颈部分,从而形成了一个沙漏形状的结构,如图 2 所示。

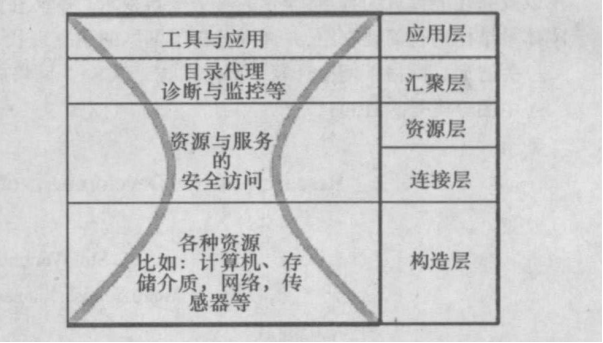


图 2 沙漏形状的五层结构<sup>[3]</sup>

2.3 开放网格服务体系结构

网络的异构性增加了其体系结构的复杂性,层出现在每个虚拟组织 (VO) 中都可能不一样。但都必须遵守传输的 QoS 特征,包括资源管理、问题决策、服务层次管理等。这些 QoS 特征需要有个定义良好的体系结构来达到所需求的服务质量,OGSA 恰好满足这样的要求。

2.3.1 OGSA 介绍

OGSA, 开放网格服务体系结构,被称为下一代网格体系结构。它是一种基于网格服务的分布式交互和计算体系结构,用来确保异构系统间的互操作性,使不同类型的系统可以相互通信、共享信息。Ian Foster, Kesselman 和 Tuecke 定义 OGSA 是支持虚拟组织来维持全体服务的创建、维持和应用<sup>[5]</sup>。

OGSA 以服务为中心,把任何资源等都抽象定义为服务。它是在 Globus 的基础上,结合最新的 Web Service 技术提出来的,它利用 Web Service 定义了 WSDL (Web Services Definition Language), WSDL 定义了服务访问的参数及其类型。所有的网格服务都是用 Globus Toolkit 工具包来构建的,所以简单地说“OGSA= 网格结构+ Web Service+ 工具包”。

2.3.2 开放网格服务基础结构 (OGSI: Open Grid Services Infrastructure)

OGSA 是一个层次结构体系,每层都有非常清晰的功能划分。其中最核心的层是 OGSI,它提供了基础构件和 OGSA 的核心平台服务,是一个标准

服务集;最高层应用和服务使用低层核心平台组件和 OGSi。

### 3 网格的研究现状及其应用

#### 3.1 国内外研究现状

网格的研究前沿主要在美国和欧洲。美国用于网格技术基础研究的经费已经超亿,其中包括由美国军方规划实施的一个宏大的网格计划,称为“全球信息网格(Global Information Grid)”,预计在 2020 年完成。Globus 是以美国阿尔贡国家实验室(ANL)为主的研发项目,全美已有 17 所大学和研究机构参与了该项目,同时 Globus 也开发了一个源码开放的 Toolkit,目前已经发行了第四版。在欧洲,英国政府也投资超亿英镑,用来研制“英国国家网格(UK National Grid)”。在亚洲,日本、印度等国家也都启动了建设国家网格的计划。我国的网格计算研究时间不长,但也取得了显著的成绩,由中国科学院牵头的“国家高性能计算环境(NHPCE)”项目和由清华大学牵头的“先进计算基础设施(ACI)——北京上海试点工程”两个网格计算项目已取得初步成果。

目前在中国已经初具规模的有“五大网格”<sup>[8]</sup>,分别是:中国国家网格(CNGrid), 863 空间信息网格,国家自然科学基金委网格(NSFC Grid), 中国教育科研网格(ChinaGrid), 上海信息网格(ShanghaiGrid)。

其中中国国家网格的试验床环境已经建设成 8 个结点:主结点有 2 个:中科院计算机网络信息中心和上海超级计算机中心,普通结点有 6 个:清华大学(北京)、中科院应用物理所(北京)、中国科学技术大学(合肥)、西安交通大学(西安)、国防科技大学(长沙)、香港大学(香港)。上海信息网格是为 2010 年上海世博会以及将上海建成国际经济、贸易、金融、航运等中心而研究建立的,它目前已经连接了上海交通大学、同济大学、上海大学、上海超级计算中心和复旦大学的超级计算机,形成了聚合计算能力 6000 亿次/秒、存储能力 4TB 的上海网格试验床。

#### 3.2 网格的应用

网格是互联网上一组新兴技术,将高速互联网、高性能计算机、大型数据库、远程设备等融为一体,为用户提供更多资源、功能和交互性,强大的网格技术被用于广泛的应用领域。

##### 3.2.1 网格在科学研究领域的应用<sup>[3]</sup>

(1) 分布式计算:目前遇到的许多科学与工程

计算问题是无法在一台超级计算机上解决的,因此需要更多的超级计算机一起来完成。网格可以把分布式的超级计算机集中起来,协同解决复杂的大规模问题。从集中计算到分布计算是网格功能的重要体现。

(2) 高吞吐率计算:高吞吐率计算关心的是几个月、一年甚至几年完成的计算量,而人们大多数情况下恰恰关心的就是一段相对较长时间内解决问题的多少,而对短期内求解问题的多少并不十分关心。对于这样的问题,可以利用 CPU 周期窃取的技术,将大量空闲计算机的计算资源集中起来,提供给对时间不太敏感的问题,作为计算资源的一种重要来源。

(3) 数据密集型计算:对于数据密集型问题,数据的采集地点、处理地点、分析与结果存放地点等往往都不同,而该类问题的求解又会产生很大的通信和计算需求,这就需要网格能力。

##### 3.2.2 网格在社会经济生活领域中的应用<sup>[3]</sup>

(1) 基于广泛信息共享的人与人交互。网格的出现可以改变原来人与人之间交流受到地理位置、交互能力、共享对象等条件的限制。

(2) 更广泛的资源贸易。计算能力闲置的机器可以共享出来,通过网格让更多的人来租用。

##### 3.2.3 网格计算在其他应用领域<sup>[2]</sup>

(1) 生命科学:在生命科学研究领域,面临着包括海量数据的分析、移动、挖掘等挑战,并且随之而产生的对于数据的安全性问题,另外对有些非标准化格式数据的查询也是急待解决的问题。网格技术的出现在一定程度上提供了解决这些海量数据问题的途径。

(2) 金融系统服务:网格计算技术为解决非常复杂的金融模型实现精确的求解提供有利条件。

(3) 高等教育:网格在不同虚拟组织间实现资源共享,并提供教育和研究机构进行交流和合作。

(4) 工程服务:包括汽车业和空间站等进行协作设计和数据分布测试。

(5) 政府部门:主要集中于提供等同的访问政府所设置的各级代理上的大量数据,为快速访问各代理,来解决政府部门间的紧急问题提供了平台。

### 4 网格的发展趋势及存在问题

目前国内外对网格的研究十分活跃。面对即将到来的第三代互联网应用,很多国家都投入了大量研究资金,希望能抓住机遇,掌握未来。网格的发展呈现出以下几个主要的趋势:

(1) 网格计算的标准化。由于网格新兴不久, 很多标准都还没有统一, 各个研究机构和研究网格的厂商都使用各自的一套机制和规范, 使得各网格很难统一起来。Globus 作为一种开放架构和开放标准基础设施, 它提供了构建网格应用所需的许多基本服务, 如安全、资源发现、资源管理、数据访问等, 目前也得到了大多数重大网格项目的支持, 逐步趋于成为一种标准。

(2) 对专用网格的研究与开发成为一个重要的研究方向。因为网格是面向具体问题的应用, 而专用网格在这方面具有独特的优势, 可以为通用网格技术提出最直接最具体的需求。

(3) Linux 将奠定网格计算的基础。网格计算和 Linux 等开放环境将改变企业的计算方式, 它将打破计算机公司加在客户身上的枷锁。

(4) 网格计算是 Internet 的未来。Internet 最初是作为一种通信平台出现的, 但随着它的发展, Internet 将演变成为一个用于计算的平台。

为了使网格计算的应用更为广泛, 实现让用户随心所欲地共享网格计算中的各种资源, 应该解决网格计算所面临的一些问题<sup>[12]</sup>: (1) 互连网的数据传输能力不足的问题; (2) 人机通信的问题; (3) 网格资源共享中的知识产权问题; (4) 网格计算的安全性; (5) 制定统一的标准, 设计一个网格操作系统。

## 5 总结

网格技术将能够让人们共享的不仅是海量数据而且是计算能力本身, 它会带来一场互连网的革命, 将改变整个计算机世界的格局, 从而给世界各行各业带来巨大的效益。网格计算汇聚了各种异构计算系统, 形成了高性能的联合计算环境, 使用网格计算可以节省购买高性能计算设备的成本和复杂计算的费用, 具有广阔的应用前景。

把握实现网格计算的关键技术以及解决网格计算目前面临的问题将成为今后网格技术研究的重要发展方向。

### 参考文献

- [1] Ian Foster, Carl Kesselman 编著. 网格计算. 金海, 袁平鹏, 石柯译(第二版)[M]. 北京: 电子工业出版社, 2004. 10
- [2] Joshy Joseph, Craig Fellenstein 著. Grid Computing. (影印版)[M]. 北京: 清华大学出版社, 2005. 1
- [3] 都志辉, 陈渝, 刘鹏 编著. 网格计算[M]. 北京: 清华大学出版社, 2002. 10
- [4] 徐志伟, 冯百明, 李伟. 网格计算技术[M]. 北京: 电子工业出版社, 2004. 5
- [5] Ian Foster, Carl Kesselman, Steven Tuecke. The anatomy of the grid: Enabling scalable virtual organizations. International J. Supercomputer Applications, 2001. 5
- [6] Ian Foster. What is the Grid? A three point checklist. [EB/OL] <http://www.-fp.mcs.anl.gov/~foster/Articles/whatisthegrid.pdf>, 2002. 7
- [7] Ian Foster, Carl Kesselman. Globus: A Metacomputing Infrastructure Toolkit. International J. Supercomputer Applications, 1997, 11.
- [8] 李明禄. 上海网格进展.[EB/OL] <http://www.chinagrid.net/dvnews/show.aspx?id=237&cid=49>, 2004. 4
- [9] 孙培德. 网格计算的研究新进展[J]. 北京: 计算机工程与应用. 2003, (16): 65~ 69
- [10] 罗作民, 张景, 李军怀等. 网格计算及其关键技术综述[J]. 北京: 计算机工程与应用 2003, (30): 18~ 20
- [11] 吴迪, 李扬, 王青海. 网格计算及其应用研究初探[J]. 南京: 电力系统通信. 2003, (6): 22~ 25
- [12] 刘忠中. 网格计算及其技术需求分析[J]. 南昌: 江西通信科技, 2003, (6): 36~ 40
- [13] 孙健, 尹晓峰, 陈光伟. 网格计算和信息共享[J]. 北京: 铁路计算机应用, 2004, 13 (4): 6~ 8
- [14] 王意洁, 肖依, 任浩等. 数据网格及其关键技术研究[J]. 北京: 计算机研究与发展, 2002, 39(8): 943~ 947
- [15] 王礼霖, 徐进. 企业网格计算应用, 北京: 微计算机应用[J]. 2004, 25 (3): 284~ 289